



**“Omixon PreciseAlign”
Geneious plugin**

User Manual

User manual for
Omixon PreciseAlign plugin
Geneious (all platforms)

March 3, 2012

Omixon Biocomputing Kft
Petzval József utca 56., Budapest, 1119 Hungary
www.omixon.com - info@omixon.com

Contents

INTRODUCTION TO THE PLUGIN.....	4
STARTING THE PLUGIN.....	5
SETTING THE PROFILES.....	6
Species Profile.....	6
Sequencer Profile.....	6
Speed Profile.....	6
Indexing Profile.....	6
ADVANCED PARAMETERS.....	7
RUNNING THE PLUGIN.....	8
ANALYZING THE RESULTS.....	9
INSTALLATION.....	10
PLUGIN INSTALLATION.....	10
UNINSTALL.....	12

Introduction to the plugin

The Omixon PreciseAlign plug-in is intended for the analysis of letter space data produced by next generation sequencing (NGS) instruments. This plug-in is designed to work with the other sequence analysis tools provided by the Geneious platform. The modules within the Omixon PreciseAlign plug-in are based on the Omixon Letter Space Toolkit.

Illumina, Ion Torrent and 454 data

The letter space module of the Omixon PreciseAlign plugin - called ORM (Omixon Read Mapper) - follows the seed-and-extend paradigm. Letter space (base space) short reads are indexed by ORM using spaced seeds and approximately mapped to a reference sequence database. ORM uses a second, much smaller seed to help to filter the approximate mappings. The underlying data structures are extremely economical for memory use, yet still provide high flexibility for trade-offs between sensitivity and specificity.

The fine alignment uses a combination of information and algorithms to produce its results, including the quality scores from the sequencer and a DNA mutation model. There are two main alignment techniques, a 'bridging' technique for smaller reads (such as Illumina short reads) where only one indel is expected, and a 'lacing' technique which allows for more indels per read, to cater for the longer Ion Torrent and Roche 454 reads.

There is special handling of repeats in the mapping, in particular tandem repeats. The tool also provides very good homopolymer error correction for Ion Torrent and Roche 454 reads.

Output

The output from the Omixon PreciseAlign plug-in is a standard Geneious 'contig' object, which can be analyzed using the tools with the Geneious 'Annotate and Predict' menu, such as 'Find Variations/SNPs', and 'Find High/Low Coverage', or exported from Geneious as a SAM file for further analysis. You can learn more about the Omixon Letter Space Toolkit [here](#).

The basic work flow supported by Geneious and the *Omixon PreciseAlign* plug-in is this:

1. Import reference genome, via File | Import or via NCBI genome search.
1. Import sequencing data (using File | Import).
2. Select both the reference and data documents and select Assembly, choosing *Omixon PreciseAlign* as the Assembler.
3. Run Annotate and Predict tools to detect variants and analyse coverage in results.

See the Geneious guides for more on steps 1, 2 and 4. This manual provides more information on step 3.

Starting the plugin

To run the Omixon PreciseAlign plugin:

- Select both a **reference genome** and **sequence** for assembly, then choose the 'Assembly' option from the main Geneious toolbar.

This opens the dialog shown in figure 1.

- Select the **Omixon PreciseAlign** option from the Assembler list in the Method section.

The screenshot shows the 'Assembly' dialog box with the following settings:

- Data:**
 - ☒ Assemble to reference: chr17
 - ☐ Assemble by: 1st part of name, separated by - (Hyphen)
 - ☐ Assemble each sequence list separately
- Method:**
 - Assembler: Omixon PreciseAlign
- Profile selection:**
 - Species: Human
 - Sequencer: Illumina
 - Speed: Normal
 - Indexing: Normal
 - Estimated memory required: 964 MB
 - Available memory: 2393 MB
 - Note: Paired reads can be set up or changed using Sequence > Set Paired Reads
- Trim Sequences:**
 - ☐ Use existing trim regions
 - ☐ Remove existing trim regions from sequences
 - ☐ Trim sequences Options
 - ☒ Do not trim
- Results:**
 - Assembly Name: brca1_0. 2 assembled to chr17
 - ☒ Save assembly report
 - ☐ Save list of unused reads
 - ☐ Save list of used reads ☐ Include mates
 - ☐ Save in sub-folder
 - ☒ Save contigs
 - ☐ Save consensus sequences Options
- Advanced:**
 - Main parameter selection:**
 - Divergence: 0.001
 - Max indel: 65
 - Default indexing: 1
 - Min alignment score: 90
 - Max alignments reported: 1
 - Max mismatches: 6
 - Other advanced parameters:**
 - Seeds: Big seed: 11111011001110101111 Small seed: 111011
 - Penalties: Indel open: 37 Indel extend: 3 Repeat open: 27 Repeat extend: 18

Buttons at the bottom: Fewer Options, OK, Cancel.

Figure 1 – Assembly options for Omixon PreciseAlign

Setting the profiles

The plug-in includes some built-in 'profiles', which can be used to easily run the mapper. These profiles are:

Species Profile

- 'Human': for mapping human data at 0.1% divergence,
- 'Bacterial 2.5': for mapping bacterial data at 2.5% divergence (appropriate for *E. coli* strains),
- 'Other': customizable profile, choosing this will make a number of extra advanced parameters available for setting - for instructions on how to set these parameters see below.

If you don't find the species you are working with within the species list you can select 'Other' and set your own species parameters

Sequencer Profile

- 'Illumina',
- 'Ion Torrent',
- '454'.

Choose which sequencer the short read data has come from. There are different algorithms and error models used for the three supported sequencers.

Speed Profile

- 'Normal': a bit slower but gives the best results
- 'Hasty': faster but the results will not be as good

ORM is both accurate (sensitive and specific) and very fast. There is also a parameter that tells the aligner how 'tenacious' (normal or hasty) to be in trying to find good mapping locations. Higher tenacity leads to slower run times but better results.

Indexing Profile

- 'Normal': index the reference sequence at every position
- 'Sparse': index the reference sequence at every fourth position
- 'Custom': choose how often to index the reference sequence, choosing this will make the 'default indexing' parameter editable in the advanced parameters section.

The default value for indexing the reference is to index at every position. This results in a larger index which uses a lot of memory, however this also gives the best results. It is possible to reduce the indexing frequency, which will save a lot of memory, however it this is at the cost of a little sensitivity and will also require a bit more processing time.

Advanced parameters

There are three advanced parameters that are always available and editable:

'Max alignments reported' - How many alignments to report in total for each read or pair. ORM will track the best alignments and if there is more than one can output this.

'Min alignment score' - The alignment step has an in-built quality filter. Reads whose scores are below this value after alignment (i.e. reads with a very low quality alignment) will automatically be discarded.

'Max indel' - The largest indel that will be allowed in an alignment.

Some of the other advanced parameters will also become available for you to change if a 'Custom' species profile is chosen, or an 'Other' indexing profile. These include the *'big seed'* and *'small seed'* and the various *alignment penalties*. The program uses two seeds, a big seed to find candidate mapping locations and a small seed to help to filter the locations. These are both gapped seeds consisting of zeroes and ones. The alignment penalties are given on the Phred scale, and are 1/10k frequency for insert/delete, and 1/1k frequency for repeats.

If you don't find the species you are working with within the species list you can select 'Other' and set your own species parameters. The most important parameter for the species is the estimated **divergence** between the sample and the reference, this must be set correctly for good results. 0.1% (human divergence) is expressed as 0.001, a bacterial divergence of 2.5% would be 0.025. **Please note that this is most important parameter - a value that is much too high or much too low can seriously affect the quality of the alignment results.**

'Default indexing' – becomes available when 'Other' is selected as the indexing profile. A value higher than 1 will reduce the memory required, at the loss of a little sensitivity within the results.

'Max mismatches' – becomes available when 'Custom' is selected as the species profile. This is a post-alignment filter, if the best alignment(s) found still have more mismatches per 100 bases than this value then they will be discarded.

Running the plugin

To start the Omixon PreciseAlign click **OK**. This plug-in will generate one kind of output, a Geneious 'contig' object:

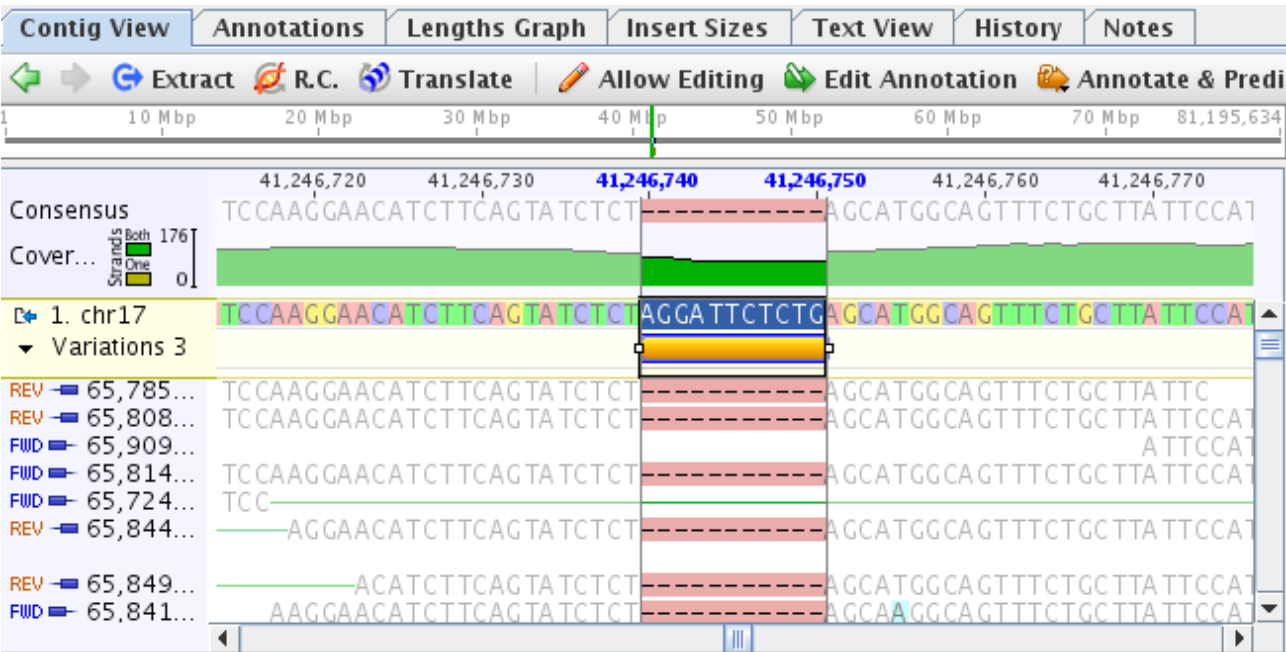


Figure 2 – Results

Analyzing the results

The various analysis tools within Geneious can be used to analyse the mapping results. For example, the variants found by Omixon PreciseAlign can also be called and displayed. Check the Geneious documentation for more info on these steps.

Variant Detection:

Annotate & Predict | Find Variations/SNPs

Coverage analysis:

Annotate & Predict | Find Low/High Coverage

As usual, the mapped reads in the 'contig' object can also be exported from Geneious in SAM format.

File | Export | Selected Documents

Installation

Plugin installation

The Omixon PreciseAlign plugin should first be downloaded via the Omixon web site:

<https://www.omixon.com/omixon/geneiousplugins.htm>

Download it now!

In order to download an evaluation version of the Omixon PreciseAlign plugin for Geneious please enter your details in the form below. We will send you an email with a link which you can use to download the tool.

Email:

* This field is required.

Full Name:

* This field is required.

Submit

Figure 3: Ordering a plugin download via the Omixon website

Once the email is received and the plugin (a .gplugin file) is downloaded, the Omixon PreciseAlign plugin is installed via the standard Geneious Plugin Preferences dialog:

Tools | Plugins from the main menu bar

The Plugins and Features preferences tab has an option to install from a file:

- **Install plugin from a gplugin file.**

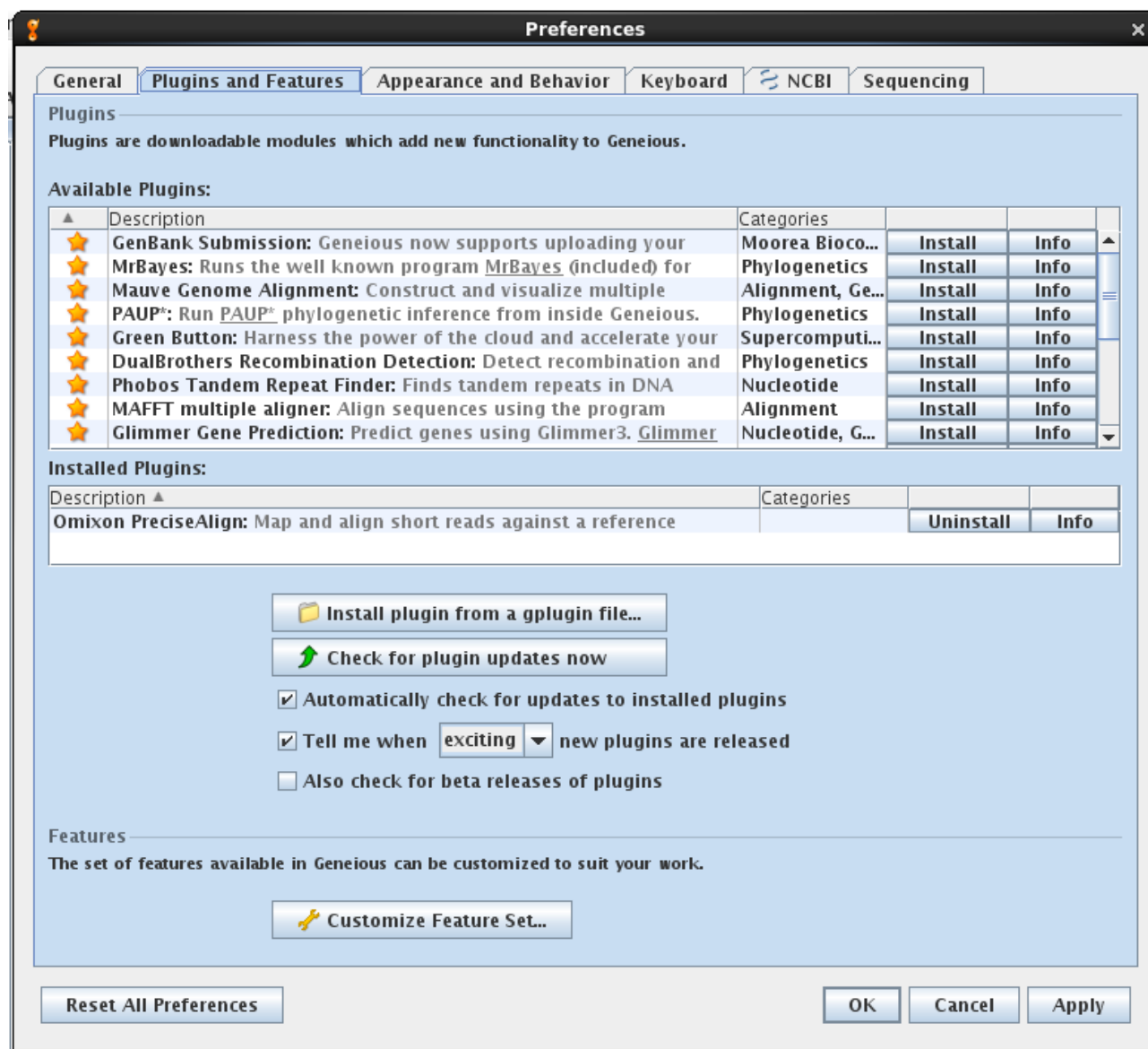


Figure 4: Installing a plugin

Use the file navigator dialog to find and select the .gplugin file downloaded earlier, **Apply** and **OK**, and installation is complete.

Uninstall

Plug-ins are uninstalled using the Geneious plugin preferences:

Tools | Plugins from the main menu bar

Select the **Uninstall** option for the plugin.